Eidgenössische Technische Hochschule Zürich
Ecole polytechnique fédérale de Zurich
Politecnico federale di Zurigo
Swiss Federal Institute of Technology Zurich

Globally optimal volume registration using DC programming

E. Kokiopoulou, D. Kressner, N. Paragios* and P. Frossard†

Research Report No. 2009-38
November 2009

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

*Laboratoire MAS, Ecole Centrale de Paris, Chatenay-Malabry, France
†Signal Processing Laboratory (LTS4), Ecole Polytechnique Fédérale de Lausanne, Switzerland

Globally Optimal Volume Registration using DC programming

E. Kokiopoulou[1], D. Kressner[1], N. Paragios[2] and P. Frossard[3]*†‡

Abstract

This paper proposes a novel approach to optimally solve rigid registration problems among volumetric images. The proposed framework exploits sparse geometric expansions for volumetric representations and DC (Difference of Convex functions) programming. We apply the SAD (sum of absolute differences) criterion to the sparse representation of the reference volume and we derive a DC decomposition of this criterion with respect to the transformation parameters. This permits to employ a cutting plane algorithm for determining the optimal relative transformation parameters of the query volume. It further enjoys theoretical guarantees for the global optimality of the obtained solution, which – to the best of our knowledge – is not offered by any other existing approach. A numerical validation demonstrates the large potential of the proposed method.

1 Introduction

Registration [1] is a fundamental problem in computer vision and in particular in medical image analysis. It is an elementary step towards bringing various volumetric data into the same reference space, which in turn permits to gather statistics and exploit similarities across subjects.

Geometric and iconic methods are often used to address this problem. Geometric methods [2] extract characteristic landmarks between two images, and then seek the optimal transformation that establishes geometric correspondences between the images. Unfortunately, such an approach may be very sensitive to the landmark extraction process. Furthermore, solving the correspondence problem between landmarks, which is a pre-step of the registration, is highly nontrivial. Often, robust EM-like methods are used for this purpose. These methods iteratively determine the optimal transformation for a set of correspondences and then improve the correspondences based on this transformation. Naturally such a method may converge to a local minimum, mostly due to erroneous correspondences.

Iconic methods [1] employ a (dis)similarity criterion on the observation space that is a function of rigid transformation parameters, which are optimized to minimize / maximize this criterion. The selection of the criterion and the optimization method are the two critical components of iconic registration. SAD, SSD, NCC, CR [3], as well as complex statistical metrics [4] in the case of multi-modal data have been considered. The optimization of the criterion is often performed using descent-like methods that are sensitive to initial conditions and do not provide guarantees on the optimality of the obtained solution. Recently the use of global optimization frameworks such as discrete MRFs was suggested [5]. However, the dimensionality of the resulting continuous search space makes its quantization quite problematic and even inefficient and therefore the results are far from being optimal.

Despite an enormous effort in the field [6], none of the existing methods can guarantee optimality of the obtained solution even in the case of volumes coming from the same modality. In this paper we propose a novel approach that estimates optimal transformation parameters. Global optimality is achieved through the expression of the objective function as a DC (difference of convex functions) decomposition and with the use of the cutting plane algorithm to estimate the optimal registration parameters.

Input volumes are sparsely represented over a redundant dictionary of geometric atoms. Using such a representation, the set of all transformations of a certain volume (which constitutes the so-called transformation manifold) admits a closed form expression with respect to the transformation parameters. This relation is used to derive a ℓ^1 criterion

*[1]Seminar for Applied Mathematics, Department of Mathematics, ETH Zurich, CH-8092 Zurich.
†[2] Laboratoire MAS, Ecole Centrale de Paris, Chatenay-Malabry, France.
‡[3]Signal Processing Laboratory (LTS4), Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne.

between the two volumes in terms of the registration parameters. Using basic theorems on DC functions [7, 8, 9], we prove that the resulting objective function admits a DC decomposition with respect to the rigid transformation parameters. Once a DC decomposition is established, a number of algorithms are available to solve the optimization problem in an efficient and robust manner [7]. In this paper, we apply the cutting plane algorithm [7, Thm 5.3] to recover the optimal registration parameters.

The rest of this paper is organized as follows. In Section 2, we briefly present the sparse geometric representations of volumes as well as the corresponding transformation manifolds. Section 3 is devoted to the definition of the registration problem, whose DC decomposition is derived in Section 4. Finally, in Section 5, we present numerical validation of our approach and some conclusions.

2 Volume transformation manifolds

In the following, we define and characterize the transformation manifold of a certain volume. For this purpose, we represent the volume by a parametric sparse model extracted from a dictionary of geometric functions. Such a geometric representation leads to a closed form expression for the transformation manifold, which is used in the computation of ℓ^1 similarity measures.

2.1 Sparse Atomic Volumetric Representations

We represent the volume of interest as a linear combination of geometric functions (usually called *atoms*), taken from a parametric and (typically overcomplete) dictionary $\mathcal{D} = \{\phi_\gamma,\ \gamma \in \Gamma\}$ spanning the input volume space. This representation generally captures the most prominent geometric features of the volume. The atoms in $\mathcal{D}$ are constructed by applying geometric transformations to a generating function denoted by ϕ. Representing the geometric transformation $\gamma \in \Gamma$ by an operator $U(\gamma)$, the parametric dictionary takes the form

$$\mathcal{D} = \{\ \phi_\gamma = U(\gamma)\phi,\ \gamma \in \Gamma\}. \tag{1}$$

In this work, a transformation $\gamma = (a, R, b) \in \Gamma$, will denote a synthesis of translations $\vec{b} \in \mathbb{R}^{3\times 1}$, anisotropic scalings $\vec{a} \in \mathbb{R}_+^{3\times 1}$ and rotations $R \in SO(3)$. The dictionary is built from three-dimensional atoms that can efficiently capture the salient geometrical features in volumetric images.

A sparse approximation of a given volume $v \in \mathbb{R}^{n_1\times n_2\times n_3}$ with atoms from the dictionary $\mathcal{D}$ can be obtained in various ways. Even if finding the sparsest approximation of v is generally a hard problem, effective sub-optimal solutions are usually sufficient to capture the salient geometric structures of a signal with only a few atoms. Such solutions are obtained, for example, by Orthogonal Matching Pursuit (OMP) [10, Sec. 9.5.3] and Tree-based Pursuit [11], to name just a few. In this work we use Tree-based Pursuit, which organizes the dictionary in a tree structure and admits significantly faster searches over the dictionary compared to OMP. Hence, this provides an effective algorithm for computing sparse volume approximations in practice. After K steps of the algorithm, the volume v is approximated by a sparse linear combination of a few atoms i.e.,

$$v = \sum_{k=1}^{K} \xi_k \phi_{\gamma_k} + r_K, \tag{2}$$

where r_K is the residual of the approximation. In what follows we will assume that r_K is negligible and can be dropped.

2.2 Characterization of transformation manifolds

The set of all geometric transformations γ applied to a certain volume v generates a manifold $\mathcal{M}$ in the high-dimensional ambient observation volume space. Each point on this manifold corresponds to a transformed version of v. In the following, we only consider transformations $\eta = (s, G, t)$ consisting of a synthesis of translations

$t = [t_x, t_y, t_z]$, *isotropic* scaling $s \in \mathbb{R}_+$ and rotations $G \in SO(3)$. Then the transformation manifold $\mathcal{M}$ can be expressed as follows:

$$\mathcal{M} = \{v(\eta) \equiv U(\eta)v, \text{ where } \eta = (s, G, t)\}. \tag{3}$$

Note that although the manifold is embedded in a high-dimensional space, its intrinsic dimension is rather small and equals the number of transformation parameters.

The transformations η form a group, namely the similitude group SIM(3) in $\mathbb{R}^3$. If (a, R, b) and (a', R', b') are two elements from SIM(3) then the group law is

$$(a, R, b) \circ (a', R', b') = (aa', RR', b + aRb'). \tag{4}$$

Using (2) and dropping the residual term r_K, it turns out that applying the transformation η to the volume v results in

$$v(\eta) = U(\eta)v = \sum_{k=1}^{K} \xi_k U(\eta)\phi_{\gamma_k} = \sum_{k=1}^{K} \xi_k \phi_{\eta\circ\gamma_k}, \tag{5}$$

where $\eta \circ \gamma_k$ is a product of transformations. In other words, the transformation is applied to each constituent atom individually, resulting in a sparse representation of the transformed volume over atoms with updated parameters. The group law (4) indeed applies [12] and can be further employed to work out the updated parameters of the transformed atoms. Equation (5) is of great importance in the proposed approach, since it expresses the manifold (3) in closed form with respect to the transformation parameters η. This is a key observation for the applicability of the DC programming methodology that is proposed in this work.

3 Rigid Registration

After having introduced sparse geometric representations and transformation manifolds, we are now ready to provide the problem formulation. We are interested in estimating the transformation between two volumes. Suppose that we are given a query volume p, and we aim to estimate the optimal transformation parameters η^* that best align v with p. We formulate the transformation estimation problem as follows

$$\eta^* = \arg \min_{\eta=(s,G,t)} f(\eta), \text{ where } f(\eta) = \|v(\eta) - p\|_1. \tag{6}$$

Here, $\|p\|_1 = \sum_{ijk} |p_{ijk}|$ denotes the ℓ^1 norm of a volume $p \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. The criterion (6) is also known as the sum of absolute differences (SAD) criterion.

Recall that $v(\eta) \in \mathcal{M}$ denotes the transformed volume v subject to a transformation $\eta = (s, G, t)$. We assume that the reference volume v has been well approximated by a sparse expansion over $\mathcal{D}$ according to (2) where r_K is negligible. Note that in the above optimization problem, only the reference volume v is expanded in the redundant basis and the query volume p is treated as is.

The optimization problem (6) is generally a non-convex nonlinear optimization problem [13] and hard to solve using traditional methods. For example, steepest descent or Newton-type methods converge only locally and may get trapped in local minima. To avoid these issues, we will show that the above objective function is a DC function with respect to the transformation parameters, i.e., it can be expressed as the difference of two convex functions.

Proposition 1 *The objective function*

$$f(\eta) = \|v(\eta) - p\|_1 = \left\| \sum_{k=1}^{K} \xi_k \phi_{\eta_k} - p \right\|_1, \tag{7}$$

where $\eta_k = \eta \circ \gamma_k$, is DC.

The proof of this proposition is given in the next section. Using Proposition 1, the optimization problem (6) can be formulated as a DC program [7, 8, 9], which can be optimally solved by exploiting the special structure of the objective function. In this paper, we employ the cutting plane method [7, Thm 5.3] to solve the DC formulation of (6). The cutting plane method is guaranteed to converge to the global minimizer. To the best of our knowledge, this is the first globally optimal algorithm that is proposed for the problem of rigid volume registration.

4 DC decomposition

In this section, we show in several steps that Proposition 1 is true. We first provide some background material on basic properties of DC functions. Using the fact that the geometric transformation of an atom ϕ_{η_k} is equivalent to a change in the coordinate system before applying $\phi(\cdot)$, we show that the transformed coordinate system $(\tilde{x}, \tilde{y}, \tilde{z})$ explicitly depends on the transformation parameters η. This is then used to show that $\tilde{x}(\eta_k)^2 + \tilde{y}(\eta_k)^2 + \tilde{z}(\eta_k)^2$ is a DC function of η, which in turn allows us to express the voxels of each atom ϕ_{η_k} in DC form. Based on the above developments, we finally obtain the DC decomposition of the objective function $\|v(\eta) - p\|_1$.

4.1 Properties of DC functions

We start with some definitions and background material about DC functions [7, 8, 9] and their properties. First, let X be a convex subset of $\mathbb{R}^n$. A function $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ is called DC on X, if there exist two convex functions $g, h : X \to \mathbb{R}$ such that f is expressed as

$$f(x) = g(x) - h(x). \tag{8}$$

A representation of the above form is called a DC decomposition of f. We present now a few properties of DC functions.

Proposition 2 ([8, Sec 4.2]) *Let $f = g - h$ and $f_i = g_i - h_i,\ i = 1\ldots, m$ be DC functions. Then the following functions are also DC:*

(a) $\sum_{i=1}^{m} \lambda_i f_i = \Big[\sum_{\{i:\lambda_i \geq 0\}} \lambda_i g_i - \sum_{\{i:\lambda_i < 0\}} \lambda_i h_i \Big] - \Big[\sum_{\{i:\lambda_i \geq 0\}} \lambda_i h_i - \sum_{\{i:\lambda_i < 0\}} \lambda_i g_i \Big].$

(b) $|f| = 2\max\{g, h\} - (g + h).$

(c) *If f_1 and f_2 are DC functions, then the product $f_1 \cdot f_2$ is DC. Moreover, if f_1 and f_2 have nonnegative convex parts, the following DC decomposition holds:*

$$\begin{aligned} f_1 \cdot f_2 \quad &= \quad \frac{1}{2}[(g_1 + g_2)^2 + (h_1 + h_2)^2] - \\ &\quad\quad \frac{1}{2}[(g_1 + h_2)^2 + (g_2 + h_1)^2]. \end{aligned} \tag{9}$$

In addition, it can be shown that the synthesis of a convex function and a DC function is again DC, which is particularly important for our further developments.

Proposition 3 *Let $f(x) : \mathbb{R}^n \to \mathbb{R}$ be DC and $q : \mathbb{R} \to \mathbb{R}$ be convex. Then,*

(a) *the composition $q(f(x))$ is DC [8, Sec 4.2].*

(b) *$q(f(x))$ has the following DC decomposition:*

$$q(f(x)) = p(x) - K[g(x) + h(x)], \tag{10}$$

where $p(x) = q(f(x)) + K[g(x) + h(x)]$ is a convex function and K is a constant satisfying $K \geq |q'(f(x))|$ [14, 15].

4.2 DC decomposition of transformed atoms

In what follows, we show that the transformed atom ϕ_{η_k} can be expressed in DC form. For notational convenience, we will drop the subscript k.

We first note that an atom in a parametric dictionary (1) is constructed by applying geometric transformations on the generating function ϕ. Applying a transformation $\gamma = (a, R, b)$ to the generating function is equivalent to transforming the coordinate system from $\{x, y, z\}$ to $\{\tilde{x}, \tilde{y}, \tilde{z}\}$ before applying $\phi(\cdot)$. More specifically, this means that an atom $\phi_\gamma = U(\gamma)\phi(x, y, z)$ coincides with $\phi(\tilde{x}, \tilde{y}, \tilde{z})$, where

$$\begin{bmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{bmatrix} = AR^\top \begin{bmatrix} x - b_x \\ y - b_y \\ z - b_z \end{bmatrix}, \tag{11}$$

and $A = \text{diag}(1/a_x, 1/a_y, 1/a_z)$.

As we have already mentioned in Sec. 2.2, a transformation η applied to ϕ_γ results in a synthesis of the two transformations η and γ. Therefore, the transformed atom $\phi_{\eta\circ\gamma}$ can be readily constructed by applying the resulting transformation $\eta \circ \gamma$ directly to the mother function as shown in the paragraph above. One should make a clear distinction between γ, which denotes the (fixed) individual transformation for each atom and η, which is the global transformation applied to the entire volume (and hence to all atoms according to (5)). The transformed coordinate system of $\phi_{\eta\circ\gamma}$ therefore depends only on η (as γ is considered fixed).

In what follows, we derive the explicit dependence between the transformed coordinate system and the transformation parameters. For this purpose, we parametrize $\eta = (s, G, t)$ using quaternions for the rotation matrix G and let (q_0, q_1, q_2, q_3) denote the quaternion parameters. This results in eight optimization variables $(s, q_0, q_1, q_2, q_3, t_x, t_y, t_z)$ for representing the transformation η.

Lemma 1 *The transformed coordinates of an atom in (7) have the form*

$$\begin{aligned} \tilde{x}(\eta) = {} & \mu_0\frac{q_0^2}{\sigma} + \mu_1\frac{q_1^2}{\sigma} + \mu_2\frac{q_2^2}{\sigma} + \mu_3\frac{q_3^2}{\sigma} + \mu_4\frac{q_1q_2}{\sigma} \\ & + \mu_5\frac{q_0q_3}{\sigma} + \mu_6\frac{q_0q_2}{\sigma} + \mu_7\frac{q_1q_3}{\sigma} + \mu_8\frac{q_2q_3}{\sigma} \\ & + \mu_9\frac{q_0q_1}{\sigma} + \mu_{10}\frac{\tau_x}{\sigma} + \mu_{11}\frac{\tau_y}{\sigma} + \mu_{12}\frac{\tau_z}{\sigma} \\ & + \mu_{13}, \end{aligned} \tag{12}$$

and similarly for $\tilde{y}$ and $\tilde{z}$ by replacing μ_i by ν_i and ζ_i, respectively. All μ_i, ν_i and ζ_i are constants depending only on the fixed atom parameters. In addition, σ as well as τ are related to s and t, respectively, by the following relations

$$\begin{aligned} \sigma &= N(q)s, \\ \tau &= \tilde{G}^\top t. \end{aligned}$$

Here, $N(q) = q_0^2 + q_1^2 + q_2^2 + q_3^2$ denotes the quaternion norm and $\tilde{G}$ denotes the (unnormalized) rotation matrix

$$\begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1q_2 + q_0q_3) & 2(q_1q_3 - q_0q_2) \\ 2(q_1q_2 - q_0q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_0q_1 + q_2q_3) \\ 2(q_0q_2 + q_1q_3) & 2(q_2q_3 - q_0q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{bmatrix}.$$

Proof. The proof is given in the appendix. □

With the change of variables suggested by the lemma above, the new optimization variables become $(\sigma, q_0, q_1, q_2, q_3, \tau_x, \tau_y, \tau_z)$. Note that we can always recover the original parameters t, s from τ, σ, and vice versa, using Lemma 1 (since the quaternion parameters are known). For notational convenience, we will assume in the following that this change of variables has been performed and continue to use η for denoting the (new) transformation parameters.

The next step in order to show that ϕ_η is DC, is to show that every constituent function in (12) is DC as well. In what follows, we provide a few lemmas towards this direction. In particular, we show that the following functions are DC: $f(q_i, \sigma) = \frac{q_i^2}{\sigma},\ i = 0, 1, 2, 3$, $f(q_i, q_j, \sigma) = \frac{q_iq_j}{\sigma},\ i, j = 0, 1, 2, 3$ and $i \neq j$, $f(\tau_x, \sigma) = \frac{\tau_x}{\sigma}$, $f(\tau_y, \sigma) = \frac{\tau_y}{\sigma}$ and $f(\tau_z, \sigma) = \frac{\tau_z}{\sigma}$.

Lemma 2 *The function $f(x, \alpha) = \frac{x}{\alpha} : \mathbb{R} \times \mathbb{R}_+^* \to \mathbb{R}$ is DC with the following DC decomposition*

$$f(x, \alpha) = \frac{x}{\alpha} = \frac{1}{2}\frac{(x+1)^2}{\alpha} - \frac{1}{2}\frac{(x^2+1)}{\alpha}. \tag{13}$$

Proof. The proof can be found in [15, Lemma 3]. □

The above lemma implies that the constituent functions $\frac{\tau_x}{\sigma}$, $\frac{\tau_y}{\sigma}$ and $\frac{\tau_z}{\sigma}$ in (12) are DC.

Lemma 3 *The function* $f(x,\alpha) = \frac{x^2}{\alpha} : \mathbb{R} \times \mathbb{R}_+^* \to \mathbb{R}$ *is convex.*

Proof. The Hessian matrix of f is

$$\nabla^2 f(x,\alpha) = \frac{1}{\alpha^3} \begin{bmatrix} 2\alpha^2 & -2x\alpha \\ -2x\alpha & 2x^2 \end{bmatrix}.$$

Observe that the term $\frac{1}{\alpha^3}$ is positive, so we only need to prove that the remaining matrix is positive semi-definite. Call λ_1 and λ_2 its eigenvalues. Then observe that its determinant is $\lambda_1\lambda_2 = 4\alpha^2 x^2 - 4x^2\alpha^2 = 0$. Thus either λ_1 or λ_2 are zero. Now, observe that the trace is $\lambda_1 + \lambda_2 = 2\alpha^2 + 2x^2 > 0$. Therefore, the Hessian matrix is positive semi-definite and f is convex.□

According to the above lemma, the constituent functions $\frac{q_i^2}{\sigma}$, $i = 0, 1, 2, 3$ in (12) are DC.

Lemma 4 *The function* $f(x,y,\alpha) = \frac{xy}{\alpha} : \mathbb{R} \times \mathbb{R} \times \mathbb{R}_+^* \to \mathbb{R}$ *is DC, with the following decomposition*

$$\frac{xy}{\alpha} = \frac{1}{2}\frac{(x+y)^2}{\alpha} - \frac{1}{2}\frac{x^2+y^2}{\alpha}. \tag{14}$$

Proof. The proof is given in the appendix. □

The above lemma implies that the functions $\frac{q_i q_j}{\sigma}$, where $i, j = 0, 1, 2, 3$ and $i \neq j$ in (12) are DC.

To summarize, we have shown the DC property of *all* constituent functions in (12). We can therefore write (12) in the more abstract form

$$\tilde{x}(\eta) = \sum_{i=0}^{13} \mu_i f_i = \sum_{i=0}^{13} \mu_i (g_i - h_i), \tag{15}$$

where $g_i - h_i$ is the DC decomposition of each function f_i. Moreover, note that each convex part g_i, h_i is nonnegative. This allows us to conclude that the transformed coordinates are DC.

Lemma 5 *The functions $\tilde{x}^2(\eta)$ (and similarly $\tilde{y}^2(\eta)$ and $\tilde{z}^2(\eta)$) introduced in Lemma 1 are DC functions of η.*

Proof. From (15) we have that $\tilde{x}^2(\eta) = \sum_{\substack{i,j=0 \\ i\neq j}}^{13} 2\mu_i\mu_j f_i f_j + \sum_{i=0}^{13} \mu_i^2 f_i^2$. Proposition 2 (c) states that the product of two DC functions (with nonnegative convex parts) is also DC. Using the results developed above, this implies that all summands in $\tilde{x}^2(\eta)$ are DC. Since the linear combination of DC functions is again DC by Proposition 2 (a) we have finally obtained that $\tilde{x}^2(\eta)$ is DC. □

Lemma 5 implies that $w(\eta) = \tilde{x}(\eta)^2 + \tilde{y}(\eta)^2 + \tilde{z}(\eta)^2$ is DC and we denote its DC decomposition by $w(\eta) = g_w(\eta) - h_w(\eta)$.

4.3 DC form of the objective function

Finally we are ready to prove the main result of our paper, namely Proposition 1, which states that the objective function of the optimization problem (6) is DC. Recall that the construction of geometric atoms by transforming the generating function is equivalent to considering the generating function on the transformed coordinates $\tilde{x}, \tilde{y}$ and $\tilde{z}$ computed above. Given these developments, it remains to show that the transformed generating functions are DC, and that the ℓ^1 distance between the transformed volume $v(\eta)$ and the query volume p is DC. We prove this for the Gaussian generating function i.e., $\phi(x,y,z) = \exp(-(x^2+y^2+z^2))$. Note that the atoms ϕ_γ are not normalized; the L^2 norm of ϕ_γ will be denoted by $\|\phi_\gamma\|$.

Proof of Proposition 1.

$$\begin{aligned} \phi_\eta &\triangleq \phi(\tilde{x}(\eta), \tilde{y}(\eta), \tilde{z}(\eta)) = \frac{e^{-(\tilde{x}(\eta)^2 + \tilde{y}(\eta)^2 + \tilde{z}(\eta)^2)}}{s\|\phi_\gamma\|} \\ &= \frac{e^{-w(\eta)}}{s\|\phi_\gamma\|} = e^{-w(\eta) - \ln s - \ln \|\phi_\gamma\|} \\ &= e^{-[w(\eta) + \ln s + \ln \|\phi_\gamma\|]} = e^{-\delta(\eta)}, \end{aligned}$$

where we have introduced the function

$$\begin{aligned} \delta(\eta) &= w(\eta) + \ln s + \ln \|\phi_\gamma\| \\ &= g_w(\eta) - h_w(\eta) + \ln s + \ln \|\phi_\gamma\|. \end{aligned} \tag{16}$$

Recall from Lemma 1 that $s = \frac{\sigma}{N(q)}$, where $N(q) = q_0^2 + q_1^2 + q_2^2 + q_3^2$, which is rewritten as

$$\ln s = \ln \sigma - \ln N(q). \tag{17}$$

Note that $\ln \sigma$ is concave. Unfortunately, $\ln N(q)$ is *not* a convex function in the quaternion parameters, and we therefore need a DC decomposition for $\ln N(q)$. We show in the appendix that

$$\begin{aligned} \ln N(q) &= \left[\ln(N(q)) + \sum_{i=0}^{3} \ln(q_i^2) \right] - \left[\sum_{i=0}^{3} \ln(q_i^2) \right] \\ &= g_{Nq}(\eta) - h_{Nq}(\eta) \end{aligned}$$

is a decomposition of $\ln N(q)$, where both components are concave. Inserting this decomposition into (17) yields $\ln s = \ln \sigma - g_{Nq}(\eta) + h_{Nq}(\eta)$. Putting all facts together, we can rewrite (16) as

$$\begin{aligned} \delta(\eta) &= [g_w(\eta) - g_{Nq}(\eta) + \ln \|\phi_\gamma\|] \\ &\quad - [h_w(\eta) - \ln \sigma - h_{Nq}(\eta)], \end{aligned} \tag{18}$$

which readily provides a DC decomposition for $\delta(\eta)$.

Next, we make use of Proposition 3 (b), which states that the synthesis of a convex function with a DC function is again DC. This shows that every voxel of ϕ_η is DC with the following decomposition: $e^{-\delta(\eta)} = [e^{-\delta(\eta)} + K(g_\delta(\eta) + h_\delta(\eta))] - [K(g_\delta(\eta) + h_\delta(\eta))]$. This holds for each atom in the sparse approximation of the volume v.

Consider now the kth atom and let $\phi_{\eta_k} = g_k(\eta) - h_k(\eta)$ denote the DC decomposition of its voxels. Next, we use once again Proposition 2 (a) to come up with the DC decomposition of $v(\eta) = \sum_{k=1}^{K} \xi_k \phi_{\eta_k}$, which reads $v(\eta) = \left[\sum_{\{k:\xi_k \geq 0\}} \xi_k g_k - \sum_{\{k:\xi_k<0\}} \xi_k h_k \right] - \left[\sum_{\{k:\xi_k \geq 0\}} \xi_k h_k - \sum_{\{k:\xi_k<0\}} \xi_k g_k \right] \equiv g_v(\eta) - h_v(\eta)$.

So far, we have shown that the transformed reference volume $v(\eta)$ is a DC decomposition of η. Since, the query volume p is fixed, the same holds for the difference volume $v(\eta) - p$. Proposition 2 (b) permits to compute the DC decomposition of the ith voxel of $|v(\eta) - p|$, which is given by $|v(\eta) - p|_i = 2\max\{g_i, h_i\} - (g_i + h_i)$, where $g_i - h_i$ is the DC decomposition of the ith voxel of $v(\eta) - p$. Finally, the objective function in (7) is DC, as it is a sum over the voxels of $|v(\eta) - p|$, which have been shown to be DC functions. □

Finally, we note that the proof above can be extended from the Gaussian generating function to other generating functions. We conclude that the objective function is a DC function, which permits the application of DC programming methods for computing the global minimizer of the optimization problem (6).

5 Numerical validation & Conclusions

Computational complexity analysis. The computational cost of the proposed method is dominated by the need for evaluating the DC decomposition of the objective function f. This scales as $O(K \cdot n_1 \cdot n_2 \cdot n_3)$, since the DC decomposition needs to be evaluated for each voxel of the K atoms of size $n_1 \times n_2 \times n_3$ whenever an evaluation of f is needed.

Numerical example. In the following, we present a numerical example to demonstrate the validity of the proposed approach. Consider a relatively simple volume v of size $16 \times 16 \times 16$, decomposed into $K = 3$ atoms. We choose a transformation η^* consisting of scaling and rotation and use it to transform v into the query volume p. Then, we use the cutting plane method [7, Thm 5.3] in order to compute an estimate $\hat{\eta}$ of η^* that aligns v with p. After a total number of 8712 iterations of the cutting plane method (with one restart at about 5000 iterations), we obtain an iterate close to the global minimizer η^*. Table 1 shows the numerical values of both the exact and the estimated transformation.

	q_0	q_1	q_2	q_3	s
η^*	-0.7303	-0.3651	0.5477	0.1826	0.8
$\hat{\eta}$	-0.7407	-0.4038	0.5166	0.1462	0.7833

Table 1: Registration results; top row shows the exact transformation parameters and the bottom row the estimated ones.

This small numerical example is intended to be a preliminary proof-of-concept, confirming the global optimality properties of the proposed approach in practice. Of course, more experimental validation is needed, especially with volumes of larger sizes, in order to explore the full potential of the proposed approach. This is exactly the focus of our current activity. We are working on a fast implementation of the methodology facilitating data-parallel processors, in particular GPUs. The obtained results will be reported in a subsequent paper.

Conclusions and Outlook We have proposed a globally optimal method for rigid registration between volumetric images by transformation parameter estimation. The proposed methodology is based on sparse volumetric representations. We have shown that under such a representation, the ℓ^1 similarity is a DC function of the transformation. This permits to solve the optimization problem using DC programming solvers that have theoretical guarantees of converging to the globally optimal solution. Finally, we have presented preliminary numerical evidence that demonstrates the large potential of the method.

The practicability of our approach crucially depends on the computational complexity, in particular the number of function evaluations needed until convergence. Future work aims at significantly reducing the complexity by adaptively interpolating the objective function and using sparse grid techniques.

References

[1] J. Maintz and M. Viergever. A survey of medical image registration. *Medical Image Analysis*, pages 1–16, 1998. 1

[2] A. Rangarajan, H. Chui, and J. Duncan. Rigid point feature registration using mutual information. *Medical Image Analysis*, pages 425–440, 1999. 1

[3] A. Roche, X. Pennec et al. Generalized correlation ratio for rigid registration of 3d ultrasound with mr images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 567 – 577, 2000. 1

[4] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. In *IEEE Transactions on Medical Imaging*, pages 187–19, 1997. 1

[5] B. Glocker, D. Zikic, N. Komodakis, N. Paragios, and N. Navab. Linear image registration through MRF optimization. In *IEEE International Symposium on Biomedical Imaging*, 2009. 1

[6] J. West, M. Fitzpatrick et al. Comparison and evaluation of retrospective intermodality brain image registration techniques. *Journal of Computer Assisted Tomography*, pages 554–568, 1997. 1

[7] R. Horst and N. V. Thoai. DC Programming: Overview. *Journal of Optimization Theory and Applications*, 102(1):1–43, October 1999. 2, 3, 4, 7

[8] R. Horst, P. M. Pardalos, and N. V. Thoai. *Introduction to Global Optimization*, volume 48 of *Nonconvex Optimization and Its Applications*. Kluwer Academic Publishers, 2nd edition, 2000. 2, 3, 4

[9] R. Horst and P. M. Pardalos. *Handbook of Global Optimization*, volume 2 of *Nonconvex Optimization and Its Applications*. Kluwer Academic Publishers, 1995. 2, 3, 4

[10] S. Mallat. *A Wavelet Tour of Signal Processing, 2nd edn.* Academic Press, 1998. 2

[11] P. Jost, P. Vandergheynst, and P. Frossard. Tree-based pursuit: Algorithm and properties. *IEEE Transactions on Signal Processing*, 54(12):4685–4697, 2006. 2

[12] R. Figueras i Ventura, P. Vandergheynst, and P. Frossard. Low rate and flexible image coding with redundant representations. *IEEE Transactions on Image Processing*, 15(3):726–739, March 2006. 3

[13] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 1999. 3

[14] S. Ellis and M. Nayakkankuppam. Phylogenetic analysis via DC programming (preprint). *Department of Mathematics and Statistics, UMBC, Baltimore, MD*, 2003. 4

[15] E. Kokiopoulou and P. Frossard. Minimum distance between pattern transformation manifolds: Algorithm and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(7):1225–1238, June 2009. 4, 6

A Proof of Lemma 1

Suppose that the atom under consideration has parameters $\gamma = (a, R, b)$, where $a = [a_x, a_y, a_z]$ and $b = [b_x, b_y, b_z]$. If we denote by $\eta = (s, G, t)$ the transformation, then according to the SIM(3) group law (4), the transformed parameters of the atom will be

$$\eta \circ \gamma = (sa, GR, t + sGb).$$

If we denote $A = \mathrm{diag}(1/a_x, 1/a_y, 1/a_z)$, then the transformed axes $[\tilde{x}, \tilde{y}, \tilde{z}]^\top$ according to (11) will be

$$\begin{aligned} \begin{pmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{pmatrix} &= \tfrac{1}{s} A R^\top G^\top \left[\begin{pmatrix} x \\ y \\ z \end{pmatrix} - t - sGb \right] \\ &= AR^\top \left[\frac{G^\top}{s} \begin{pmatrix} x \\ y \\ z \end{pmatrix} - \frac{G^\top}{s} t - b \right]. \end{aligned} \tag{19}$$

In the above equation we have used the fact that G is a rotation matrix i.e., $G^\top G = I$.

We use quaternions to parameterize the unknown rotation matrix G. Consider a quaternion $q = q_0 + iq_1 + jq_2 + kq_3$ with norm $N(q) = q_0^2 + q_1^2 + q_2^2 + q_3^2$. Then the corresponding rotation matrix takes the form

$$G = \frac{1}{N(q)} \tilde{G},$$

where

$$\tilde{G} = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1 q_2 + q_0 q_3) & 2(q_1 q_3 - q_0 q_2) \\ 2(q_1 q_2 - q_0 q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_0 q_1 + q_2 q_3) \\ 2(q_0 q_2 + q_1 q_3) & 2(q_2 q_3 - q_0 q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{bmatrix} \tag{20}$$

denotes the unnormalized rotation matrix. Inserting this representaion into (19) gives

$$\begin{pmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{pmatrix} = AR^\top \left[\frac{\tilde{G}^\top}{N(q)s} \begin{pmatrix} x \\ y \\ z \end{pmatrix} - \frac{\tilde{G}^\top}{N(q)s} t - b \right]. \tag{21}$$

Defining

$$\sigma = N(q)s, \tag{22}$$

$$\tau = \tilde{G}^\top t, \tag{23}$$

the original set of optimization variables $(s, q_0, q_1, q_2, q_3, t_x, t_y, t_z)$ becomes $(\sigma, q_0, q_1, q_2, q_3, \tau_x, \tau_y, \tau_z)$. Note that these two variable representations are equivalent and one may switch from the first one to the second and vice versa via the use of equations (22) and (23). In what follows we use the second representation and rewrite (21) as

$$\begin{pmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{pmatrix} = AR^\top \left[\frac{1}{\sigma} \tilde{G}^\top \begin{pmatrix} x \\ y \\ z \end{pmatrix} - \frac{1}{\sigma} \tau - b \right]. \tag{24}$$

Recall that in the above, A, R, b, x, y and z are constant. In particular, the matrix $AR^\top$ is constant and its entries are denoted as follows:

$$AR^\top = \begin{bmatrix} \rho_1 & \rho_2 & \rho_3 \\ \rho_4 & \rho_5 & \rho_6 \\ \rho_7 & \rho_8 & \rho_9 \end{bmatrix}.$$

The right hand side of (24) thus takes the form

$$\begin{bmatrix} \rho_1 & \rho_2 & \rho_3 \\ \rho_4 & \rho_5 & \rho_6 \\ \rho_7 & \rho_8 & \rho_9 \end{bmatrix} \frac{1}{\sigma} \left[\tilde{G}^\top \begin{pmatrix} x \\ y \\ z \end{pmatrix} - \tau \right] - \begin{pmatrix} c_x \\ c_y \\ c_z \end{pmatrix}, \tag{25}$$

where $[c_x, c_y, c_z]^\top = AR^\top b$.

Next, we will use (20) to compute the explicit dependence of $\tilde{x}$ on the optimization variables. Note that

$$\tilde{G}^\top \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{bmatrix} x(q_0^2 + q_1^2 - q_2^2 - q_3^2) + 2y(q_1 q_2 - q_0 q_3) + 2z(q_0 q_2 + q_1 q_3) \\ 2x(q_1 q_2 + q_0 q_3) + y(q_0^2 - q_1^2 + q_2^2 - q_3^2) + 2z(q_2 q_3 - q_0 q_1) \\ 2x(q_1 q_3 - q_0 q_2) + 2y(q_0 q_1 + q_2 q_3) + z(q_0^2 - q_1^2 - q_2^2 + q_3^2) \end{bmatrix}.$$

Putting all the above facts together one finally obtains, after some straightforward algebraic manipulation,

$$\begin{aligned} \tilde{x} \;=\; & \mu_0 \frac{q_0^2}{\sigma} + \mu_1 \frac{q_1^2}{\sigma} + \mu_2 \frac{q_2^2}{\sigma} + \mu_3 \frac{q_3^2}{\sigma} + \mu_4 \frac{q_1 q_2}{\sigma} + \mu_5 \frac{q_0 q_3}{\sigma} \\ & + \mu_6 \frac{q_0 q_2}{\sigma} + \mu_7 \frac{q_1 q_3}{\sigma} + \mu_8 \frac{q_2 q_3}{\sigma} + \mu_9 \frac{q_0 q_1}{\sigma} + \mu_{10} \frac{\tau_x}{\sigma} \\ & + \mu_{11} \frac{\tau_y}{\sigma} + \mu_{12} \frac{\tau_z}{\sigma} + \mu_{13}, \end{aligned} \tag{26}$$

where

$$\begin{aligned} \mu_0 &= \rho_1 x + \rho_2 y + \rho_3 z \\ \mu_1 &= \rho_1 x - \rho_2 y - \rho_3 z \\ \mu_2 &= -\rho_1 x + \rho_2 y - \rho_3 z \\ \mu_3 &= -\rho_1 x - \rho_2 y + \rho_3 z \\ \mu_4 &= 2\rho_1 y + 2\rho_2 x \\ \mu_5 &= -2\rho_1 y + 2\rho_2 x \\ \mu_6 &= 2\rho_1 z - 2\rho_3 x \\ \mu_7 &= 2\rho_1 z + 2\rho_3 x \\ \mu_8 &= 2\rho_2 z + 2\rho_3 y \\ \mu_9 &= -2\rho_2 z + 2\rho_3 y \\ \mu_{10} &= -\rho_1 \\ \mu_{11} &= -\rho_2 \\ \mu_{12} &= -\rho_3 \\ \mu_{13} &= -c_x. \end{aligned}$$

Observe that all μ_i are constant. This concludes the proof for $\tilde{x}$. The derivation for $\tilde{y}$ and $\tilde{z}$ is similar and therefore omitted. □

B Proof of Lemma 4

We need to show that the two components in (14) are convex. We start with the function $f(x, y, a) = \frac{(x+y)^2}{a},\ a > 0$. The Hessian matrix $\nabla^2 f(x, y, a)$ is

$$a^3 \begin{bmatrix} 2a^2 & 2a^2 & -2a(x+y) \\ 2a^2 & 2a^2 & -2a(x+y) \\ -2a(x+y) & -2a(x+y) & (x+y)^2 \end{bmatrix}.$$

Now consider a vector $v = [v_1, v_2, v_3]^\top$ and observe that

$$\begin{aligned} v^\top \nabla^2 f(x,y,a) v &= 2a^2 v_1^2 + 2a^2 v_2^2 + 2(x+y)^2 v_3^2 \\ &\quad +4a^2 v_1 v_2 - 4a(x+y) v_1 v_3 \\ &\quad -4a(x+y) v_2 v_3 \\ &= 2(a v_1 + a v_2 - (x+y) v_3)^2 \geq 0. \end{aligned}$$

Hence the first component is convex.

Considering now the second component $f(x,y,a) = \frac{x^2+y^2}{a},\ a > 0$, the Hessian matrix is

$$\nabla^2 f(x,y,a) = a^3 \begin{bmatrix} 2a^2 & 0 & -2xa \\ 0 & 2a^2 & -2ya \\ -2xa & -2ya & 2(x^2+y^2) \end{bmatrix}.$$

Similarly as above, consider a vector $v = [v_1, v_2, v_3]^\top$ and observe that

$$\begin{aligned} v^\top \nabla^2 f(x,y,a) v &= 2[a^2 v_1^2 + a^2 v_2^2 \\ &\quad +(x^2+y^2) v_3^2 - 2xa v_1 v_3 \\ &\quad -2ya v_2 v_3] \\ &= 2(a v_1 - x v_3)^2 + 2(a v_2 - y v_3)^2 \geq 0, \end{aligned}$$

which shows that the second part is also convex. $\square$

C Decomposition of $\ln(x^2 + y^2 + z^2 + w^2)$

Let $N = x^2 + y^2 + z^2 + w^2$. We will show that the following decomposition

$$\begin{aligned} \ln N &= \left[\ln(N) + \ln(x^2) + \ln(y^2) + \ln(z^2) + \ln(w^2)\right] - \\ &\quad \left[\ln(x^2) + \ln(y^2) + \ln(z^2) + \ln(w^2)\right], \end{aligned} \tag{27}$$

has concave components.

Proof. The second component is concave as it consists of a sum of concave scalar functions. We now focus on the first part, whose Hessian matrix is

$$H = \frac{1}{N^2} \begin{bmatrix} \zeta(x) & -4xy & -4xz & -4xw \\ -4xy & \zeta(y) & -4yz & -4yw \\ -4xz & -4yz & \zeta(z) & -4zw \\ -4xw & -4yw & -4zw & \zeta(w) \end{bmatrix},$$

where we have introduced the function $\zeta(x) = \frac{2x^2N - 2N^2 - 4x^4}{x^2}$ for notational convenience. Consider a vector $v = [v_1, v_2, v_3, v_4]^\top$. After factoring out the term $\frac{1}{x^2y^2z^2w^2}$ and some algebraic manipulation, it holds that

$$
\begin{aligned}
v^\top H v \quad = \quad & -2v_1^2w^6y^2z^2 - 2v_1^2w^4x^2y^2z^2 - 4v_1^2w^4y^4z^2 \\
& -4v_1^2w^4y^2z^4 - 4v_1^2w^2x^4y^2z^2 - 2v_1^2w^2x^2y^4z^2 \\
& -2v_1^2w^2x^2y^2z^4 - 2v_1^2w^2y^6z^2 - 4v_1^2w^2y^4z^4 \\
& -2v_1^2w^2y^2z^6 - 8v_1v_2w^2x^3y^3z^2 - 8v_1v_3w^2x^3y^2z^3 \\
& -8v_1v_4w^3x^3y^2z^2 - 2v_2^2w^6x^2z^2 - 4v_2^2w^4x^4z^2 \\
& -2v_2^2w^4x^2y^2z^2 - 4v_2^2w^4x^2z^4 - 2v_2^2w^2x^6z^2 \\
& -2v_2^2w^2x^4y^2z^2 - 4v_2^2w^2x^4z^4 - 4v_2^2w^2x^2y^4z^2 \\
& -2v_2^2w^2x^2y^2z^4 - 2v_2^2w^2x^2z^6 - 8v_2v_3w^2x^2y^3z^3 \\
& -8v_2v_4w^3x^2y^3z^2 - 2v_3^2w^6x^2y^2 - 4v_3^2w^4x^4y^2 \\
& -4v_3^2w^4x^2y^4 - 2v_3^2w^4x^2y^2z^2 - 2v_3^2w^2x^6y^2 \\
& -4v_3^2w^2x^4y^4 - 2v_3^2w^2x^4y^2z^2 - 2v_3^2w^2x^2y^6 \\
& -2v_3^2w^2x^2y^4z^2 - 4v_3^2w^2x^2y^2z^4 - 8v_3v_4w^3x^2y^2z^3 \\
& -4v_4^2w^4x^2y^2z^2 - 2v_4^2w^2x^4y^2z^2 - 2v_4^2w^2x^2y^4z^2 \\
& -2v_4^2w^2x^2y^2z^4 - 2v_4^2x^6y^2z^2 - 4v_4^2x^4y^4z^2 \\
& -4v_4^2x^4y^2z^4 - 2v_4^2x^2y^6z^2 - 4v_4^2x^2y^4z^4 \\
& -2v_4^2x^2y^2z^6 \\
= \quad & -2w^2z^2(v_1y^3 + v_2x^3)^2 - 2x^2y^2z^2w^2(v_1y + v_2x)^2 \\
& -2w^2y^2(v_1z^3 + v_3x^3)^2 - 2x^2y^2z^2w^2(v_1z + v_3x)^2 \\
& -2y^2z^2(v_1w^3 + v_4x^3)^2 - 2x^2y^2z^2w^2(v_1w + v_4x)^2 \\
& -2w^2x^2(v_2z^3 + v_3y^3)^2 - 2x^2y^2z^2w^2(v_2z + v_3y)^2 \\
& -2x^2z^2(v_2w^3 + v_4y^3)^2 - 2x^2y^2z^2w^2(v_2w + v_4y)^2 \\
& -2x^2y^2(v_3w^3 + v_4z^3)^2 - 2x^2y^2z^2w^2(v_3w + v_4z)^2 \\
& -4v_1^2w^4y^4z^2 - 4v_1^2w^4y^2z^4 - 4v_1^2w^2x^4y^2z^2 \\
& -4v_1^2w^2y^4z^4 - 4v_2^2w^4x^4z^2 - 4v_2^2w^4x^2z^4 \\
& -4v_2^2w^2x^4z^4 - 4v_2^2w^2x^2y^4z^2 - 4v_3^2w^4x^4y^2 \\
& -4v_3^2w^4x^2y^4 - 4v_3^2w^2x^4y^4 - 4v_3^2w^2x^2y^2z^4 \\
& -4v_4^2w^4x^2y^2z^2 - 4v_4^2x^4y^4z^2 - 4v_4^2x^4y^2z^4 \\
& -4v_4^2x^2y^4z^4 \leq 0.
\end{aligned} \tag{28}
$$

Hence, the first part of the decomposition is concave. □

Research Reports

No. Authors/Title

09-38 *E. Kokiopoulou, D. Kressner, N. Paragios, P. Frossard*
Globally optimal volume registration using DC programming

09-37 *F.G. Fuchs, A.D. McMurray, S. Mishra, N.H. Risebrom, K. Waagan*
Approximate Riemann solvers and stable high-order finite volume schemes for multi-dimensional ideal MHD

09-36 *Ph. LeFloch, S. Mishra*
Kinetic functions in magnetohydrodynamics with resistivity and hall effects

09-35 *U.S. Fjordholm, S. Mishra*
Vorticity preserving finite volume schemes for the shallow water equations

09-34 *S. Mishra, E. Tadmor*
Potential based constraint preserving genuinely multi-dimensional schemes for systems of conservation laws

09-33 *S. Mishra, E. Tadmor*
Constraint preserving schemes using potential-based fluxes.
III. Genuinely multi-dimensional central schemes for for MHD equations

09-32 *S. Mishra, E. Tadmor*
Constraint preserving schemes using potential-based fluxes.
II. Genuinely multi-dimensional central schemes for systems of conservation laws

09-31 *S. Mishra, E. Tadmor*
Constraint preserving schemes using potential-based fluxes.
I. Multidimensional transport equations

09-30 *D. Braess, S. Sauter, C. Schwab*
On the justification of plate models

09-29 *D. Schötzau, C. Schwab, T. Wihler*
hp-dGFEM for second-order elliptic problems in polyhedra.
II: Exponential convergence

09-28 *D. Schötzau, C. Schwab, T. Wihler*
hp-dGFEM for second-order elliptic problems in polyhedra.
I: Stability and quasioptimality on geometric meshes

09-27 *A. Moiola, R. Hiptmair, I. Perugia*
Approximation by plane waves

09-26 *M. Karow, E. Kokiopoulou, D. Kressner*
On the computation of structured singular values and pseudospectra